# ASSESING THE NUMBER OF LINEAR COMPONENTS IN A GENERAL REGRESSION PROBLEM

Santiago Velilla

97-20

WORKING PAPERS

Universidad Carlos III de Madrid

Working Paper 97-20
Statistics and Econometrics Series 11
March 1997

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

# ASSESSING THE NUMBER OF LINEAR COMPONENTS IN A GENERAL REGRESSION PROBLEM

Santiago Velilla*

Abstract

This paper presents a method for testing the dimension of a general regression problem without assuming any specific model for the distribution of the regressors. Examples of application and comparisons with existing methods are studied.

Key Words

Asymptotic tests, dimension reduction, eigenvalues, sliced inverse regression

*Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, e-mail: savece@est-econ.uc3m.es.

# Assessing the Number of Linear Components in a General Regression Problem

Santiago Velilla*

**Abstract**

This paper presents a method for testing the dimension of a general regression problem without assuming any specific model for the distribution of the regressors. Examples of application and comparisons with existing methods are studied.

**Keywords and phrases:** asymptotic tests, dimension reduction, eigenvalues, sliced inverse regression.

## 1. INTRODUCTION

Consider a scalar response $y$ that is to be explained in terms of a $p$ dimensional set of carriers $\mathbf{x} = (x_1, \ldots, x_p)'$. To study the dependence on $\mathbf{x}$ of the conditional distribution $y \mid \mathbf{x}$, Li (1991) introduces the general regression model

$$y = f(\beta_1'\mathbf{x}, \ldots, \beta_k'\mathbf{x}, \epsilon), \tag{1.1}$$

where $f(.)$ is an arbitrary function on $\mathbf{R}^{k+1}$, $\beta_1, \ldots, \beta_k$ are unknown and linearly independent $p \times 1$ vectors, and $\epsilon$ is a random error independent of $\mathbf{x}$. Under (1.1),

*Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, 28903 - Getafe, Madrid, Spain. Research partailly supported by *DGICYT* Grant *PB*93 – 0232 (*Spain*).

the conditional distributions $y \mid \mathbf{x}$ and $y \mid B'\mathbf{x}$ are the same, where $B = (\beta_1, ..., \beta_k)$. Moreover, $y$ and $\mathbf{x}$ are conditionally independent given $B'\mathbf{x}$ (Dawid 1979; Cook 1994) so $\mathbf{x}$ can be replaced by $B'\mathbf{x}$ without loss of information. When $p$ is large and $k \ll p$, this property characterizes (1.1) as a useful tool for dimension reduction. Notice that the conditional distributions $y \mid B'\mathbf{x}$ and $y \mid A'\mathbf{x}$ are identical for any $p \times k$ matrix $A$ such that its columns span $C(B)$, the column space of $B$. Therefore, unless some additional restrictions are imposed on the unspecified function $f(.)$ of (1.1), $C(B)$ is identifiable but the vectors $\beta_1$, ..., $\beta_k$ are not. $C(B)$ is called the effective dimension reduction (e.d.r) space in Li (1991). Conditions for the uniqueness of $C(B)$ are given in Cook (1994). Borrowing terminology from Cook and Weisberg (1994, $p.$ xvii), $k = \dim C(B)$ is the structural dimension of the regression.

Although in most regression problems $k \leq 2$, $k$ is, in principle, an unknown parameter of model (1.1) and some statistical techniques have been developed for making inferences on the number of linear components describing the relation between $y$ and $\mathbf{x}$. Existing methods are reviewed in section 2. Some of them depend on specific assumptions on the distribution of $\mathbf{x}$ such as, for example, normality or elliptical symmetry. This paper introduces a new testing procedure that uses only general aspects of the distribution of the regressors. Main results are exposed in section 3. Section 4 presents simulations and examples of application and section 5 gives some final comments.

## 2. BACKGROUND AND MOTIVATION

The inverse regression curve $E(\mathbf{x} \mid y)$ plays a central role in the analysis of (1.1). Suppose that $\mathbf{x}$ has mean $\mu$ and dispersion matrix $\Sigma$, and define $\mathbf{z} = \Sigma^{-1/2}(\mathbf{x}-\mu)$. Assuming that for every fixed $p \times 1$ vector $b$ the conditional expectation of $b'\mathbf{x}$ given $B'\mathbf{x}$ is of the form

$$E(b'\mathbf{x} \mid B'\mathbf{x}) = c_0 + c_1(\beta_1'\mathbf{x}) + ... + c_k(\beta_k'\mathbf{x}), \tag{2.1}$$

for some constants $c_0$, ..., $c_k$, theorem 3.1 in Li (1991, p. 349) proofs that, under (1.1), the standardized inverse regression curve $E(\mathbf{z} \mid y)$ is contained in $C(\Sigma^{1/2}B) = C(\eta_1, ..., \eta_k)$, where $\eta_j = \Sigma^{1/2}\beta_j$ is the $jth$ standardized e.d.r. direction. This is an important result that allows to avoid the curse of dimensionality when $p$ is large. The analysis is based on the $p$ one dimensional inverse regression curves $E(x_j \mid y)$ rather than on the complex forward regression $E(y|\ \mathbf{x})$.

### 2.1 Sliced inverse regression and other methods for finding dimensionality

The matrix

$$V = V[E(\mathbf{z} \mid y)] \tag{2.2}$$

is degenerate in any direction orthogonal to the standardized e.d.r. directions $\eta_1$, ..., $\eta_k$. Li (1991) suggests the algorithm $SIR$ (sliced inverse regression) for estimating $\beta_1$, ..., $\beta_k$ given i.i.d. observations $(y_i,\ x_i^{'})$, $i = 1,\ ...,\ n$. Divide the range of the data into $H$ nonoverlapping slices according to the values of the response $y$, and let $I_h$ be the $hth$ generical slice and $n_h$ the number of responses in $I_h$, $h = 1,\ ...,\ H$. $SIR$ estimates $V$ by the weighted $p \times p$ covariance matrix

$$\widehat{V}_n = \sum_{h=1}^{H} \widehat{p}_h \widehat{m}_h \widehat{m}_h^{'}, \tag{2.3}$$

where $\widehat{p}_h = n_h/n$, $\widehat{m}_h$ is the average of the standardized predictors $z_i = S^{-1/2}(x_i - \overline{x})$ such that $y_i \in I_h$, and $S$ and $\overline{x}$ are, respectively, the sample dispersion matrix and mean vector of the $x_i$. Let $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq\ ...\ \geq \widehat{\lambda}_p$ be the ordered eigenvalues of $\widehat{V}_n$ and $\widehat{\eta}_1$, ..., $\widehat{\eta}_k$ the eigenvectors corresponding to the $k$ largest eigenvalues. The estimates of the e.d.r. directions are $\widehat{\beta}_j = S^{-1/2}\widehat{\eta}_j$. Li (1991) also gives a test of dimensionality. Assuming (1.1) and $\mathbf{x} \sim N_p(\mu, \Sigma)$, the statistic

$$L_k = n \sum_{i=k+1}^{p} \widehat{\lambda}_i \tag{2.4}$$

has an asymptotic $\chi^2_{(p-k)(H-k-1)}$ distribution. The null that the regression has at most $k$ components versus the alternative that the regression has at least $k+1$ components

is rejected for large values of $L_k$. $SIR$ is studied from a theoretical viewpoint in Duan and Li (1991), expanding previous work in Li and Duan (1989).

Schott (1994) relaxes the assumption of normality and develops a testing method assuming only an elliptically symmetric distribution for $\mathbf{x}$. Graphical techniques for finding dimension are discussed in Cook and Weisberg (1994, chap. 8). See also Cook and Wetzel (1993). Formal inference methods and plotting procedures are seen as complementary analytical tools, the findings of one approach being confirmed by the other and vice versa.

$SIR$ is not effective in symmetric situations. For example, if $y = f(\beta'\mathbf{x}) + \epsilon$ and both $\mathbf{x}$ and $f$ are symmetric around the origin, then $(y, \mathbf{x}') \stackrel{D}{=} (y, -\mathbf{x}')$ and the inverse regression curve $E(\mathbf{x} \mid y) = 0$ is contained in a trivial subspace of $C(\beta)$. Some remedial actions have been suggested using second moment inverse regression methods. Cook and Weisberg (1991) propose assessing dimension with the larger eigenvalues of the sample version of $SAVE = \sum_h [I_p - var(\mathbf{z} \mid y \in I_h)]^2$. $SAVE$ is an acronym for sliced average variance estimate. Li (1991) introduces a class of methods, labeled $SIRII$, that use the inverse conditional curve $cov(\mathbf{x} \mid y)$. Finally, Li (1992) provides another method based on the notion of principal Hessian directions ($pHd$) that improves the performance of $SIR$ in symmetric situations.

### 2.2 An example: air quality data

This example, taken from Cook and Weisberg (1994, sec. 8.4), illustrates some of the issues that appear when testing for dimension. The data give air quality readings for $n = 111$ nearly consecutive days in the New York City area in 1973. The response variable is *Ozone*, the Ozone concentration in parts per billion. Three predictors are considered: *SolR*, solar radiation in Langleys, *Wind*, the wind speed in miles per hour, and *Temp*, the temperature in degrees Fahrenheit. The scatter plot matrix of figure 1 and other outputs presented below are obtained using the $R - code$ (Cook and Weisberg 1994).

*Figure* 1

Notice that the plots of $SolR$ versus $Wind$ and $SolR$ versus $Temp$ show a nonelliptical pattern so direct application of Li's test statistic (2.4) is perhaps questionable. Table 1 presents two different analysis of dimension based on (2.4) corresponding, respectively, to choosing $H = 19$ and $H = 26$ slices according to the values of $Ozone$. The first analysis indicates one dimensional structure while the seconds suggests 2D structure.

*Table* 1

Cook and Weisberg (1994) give a very detailed analysis of this data set and attribute the anomaly observed in table 1 to the role, in choosing the slices, of both the seven largest and nine smallest values of $Ozone$.

## 3. A GENERAL TEST OF DIMENSION

As the example above motivates, there is a need of building a method for testing for dimension that does not depend on specific assumptions on the distribution of the regressors. The role of the number of slices $H$ as a tuning constant of a procedure like (2.4) should be also clarified. This section presents a proposal of a test of dimension valid for a wide class of distributions of $\mathbf{x}$ verifying condition (2.1). Hall and Li (1993) have shown that, in general, this condition will be satisfied in problems with a large number of regressors. Notice that (2.1) is implied by but is not equivalent to elliptical symmetry.

### 3.1 Preliminaries

Order the data $(y_i, x_i')$, $i = 1, ..., n$, according to the increasing values $y_{(1)} \leq y_{(2)} \leq ... \leq y_{(n)}$ of $y$ and arrange the responses in slices with the same amount $c \geq 2$ of observations. The notation $x_{(i)}$ will be used for the values of the predictors corresponding to $y_{(i)}$. Yang (1977) calls $x_{(i)}$ the concomitant value of the order statistic $y_{(i)}$. Before presenting the test of dimension, an important asymptotic result

is necessary. Put

$$\Lambda = E[V(\mathbf{x} \mid y)], \tag{3.1}$$

and

$$\hat{\Lambda}_n = \sum_{h=1}^{H} \frac{c}{n} \hat{S}_h, \tag{3.2}$$

where $\hat{S}_h = \frac{1}{c-1} \sum_{j=1}^{c} [x_{(h,j)} - \hat{\mu}_{(h)}][x_{(h,j)} - \hat{\mu}_{(h)}]'$, $x_{(h,j)}$ is the concomitant value of the $jth$ ordered response in the $hth$ slice and $\hat{\mu}_{(h)} = \sum_{j=1}^{c} x_{(h,j)}/c$. The notation in (3.2) should be understood to indicate that the number of slices is $H = H_n = [n/c]$ when $n/c$ is an integer, and $H = H_n = [n/c] + 1$ otherwise, where $[.]$ denotes integer part. The number of observations in the last slice is then $n - c[n/c]$. Assuming certain regularity conditions on the moments of $\mathbf{x}$, on the inverse regression curve $E(\mathbf{x} \mid y)$ and on $V(y) = V(\mathbf{u} \mid y)$, where $\mathbf{u} = \mathbf{x} - E(\mathbf{x} \mid y)$, theorem 1 in Zhu and Ng (1995, p. 730) establishes

$$n^{1/2}(\hat{\Lambda}_n - \Lambda) \xrightarrow{D} W, \tag{3.3}$$

where $W$ is a random $p \times p$ symmetric matrix such that $vech(W) \sim N_{p(p+1)/2}(0, \Gamma)$ and $\Gamma = V[vech(\mathbf{u}\mathbf{u}')] + \frac{2}{c-1} E[vech(V(y))\, vech(V(y))']$. The convergence (3.3) extends a previous result of Hsing and Carroll (1992) for the case $c = 2$ and can be used as a building block for a general test of dimension.

### 3.2 A test for the number of components

From the identity $\Sigma = V(\mathbf{x}) = E[V(\mathbf{x} \mid y)] + V[E(\mathbf{x} \mid y)] = \Lambda + V[E(\mathbf{x} \mid y)$, it follows

$$I_p = \Sigma^{-1/2} \Lambda \Sigma^{-1/2} + V, \tag{3.4}$$

where $V$ is as in (2.2). From (3.4),

$$\begin{aligned} & n^{1/2}[(I_p - S^{-1/2} \hat{\Lambda}_n S^{-1/2}) - V] \\ = \; & n^{1/2}[(I_p - S^{-1/2} \hat{\Lambda}_n S^{-1/2}) - (I_p - \Sigma^{-1/2} \Lambda \Sigma^{-1/2})] \\ = \; & -n^{1/2}[S^{-1/2} \hat{\Lambda}_n S^{-1/2} - \Sigma^{-1/2} \Lambda \Sigma^{-1/2}], \end{aligned} \tag{3.5}$$

so it can be conjectured, using (3.3), that the matrix (3.5) is asymptotically normal. Let $\lambda_1 \geq \lambda_2 \geq \ ... \ \geq \lambda_p \geq 0$ be the ordered eigenvalues of $V$ and, correspondingly, let $\hat{\alpha}_1 \geq \hat{\alpha}_2 \geq \ ... \ \geq \hat{\alpha}_p$ be the ordered eigenvalues of $I_p - S^{-1/2}\hat{\Lambda}_n S^{-1/2}$. If the rank of $V$ is $k$, then $\lambda_1 \geq \lambda_2 \geq \ ... \ \geq \lambda_k > 0 = \lambda_{k+1} = \lambda_{k+2} = \ ... \ = \lambda_p$. The result below follows.

**Theorem 1** *Under the regularity conditions of theorem 1 in Zhu and Ng (1995, p. 730) and the null $H_k : rank(V) = k$, the statistic*

$$S_k = n^{1/2} \sum_{j=k+1}^{p} \hat{\alpha}_j \xrightarrow{D} N(0, \sigma_k^2), \tag{3.6}$$

*where $\sigma_k^2$ is a positive constant.*

**Proof.** The idea is to establish first joint asymptotic normality of the $p \times 2p$ matrix $n^{1/2}[(S, \hat{\Lambda}_n) - (\Sigma, \Lambda)]$ using (3.3). Application of the delta method solves then the asymptotics for (3.5). The limit distribution of the random vector $n^{1/2}(\hat{\alpha}_{k+1}, \hat{\alpha}_{k+2}, ..., \hat{\alpha}_p)'$ can be derived using theorem 3.2 in Eaton and Tyler (1991, p. 265). The rest follows from the Cramér–Wold device. Details are in appendix A. ■

Theorem 1 leads to a test of dimension. Starting with $k = 0$, the null $H_k$ : $rank(V) = k$ is rejected for large values of the criterion

$$G_k = S_k / \hat{\sigma}_k, \tag{3.7}$$

where $S_k$ is as in (3.6) and $\hat{\sigma}_k$ is a proper estimate of $\sigma_k$. The construction of $\hat{\sigma}_k$ is discussed in subsection 3.3 below. Approximate p-values are computed with reference to the standard $N(0, 1)$ distribution. If $H_k$ is rejected, $k$ is increased by one and the procedure continues until reaching an integer $K$ such that $G_K$ is not significant. Recalling (2.2) and theorem 3.1 of Li (1991), the dimension of a regression model of the form (1.1) is declared as being at least $K$.

### 3.3 Construction of $\hat{\sigma}_k$

By expression (A.14) of the proof of theorem 1 given in appendix A, the constant $\sigma_k^2$ in (3.6) is a quadratic form

$$\sigma_k^2 = u_k' M u_k, \tag{3.8}$$

where

$$u_k = vec(C_2 C_2'), \tag{3.8a}$$

is a $p^2 \times 1$ vector that depends on the $p \times (p-k)$ matrix $C_2 = (\gamma_{k+1}, \ ..., \ \gamma_p)$ formed by the orthonormalized eigenvectors associated to the $p-k$ smallest eigenvalues of $V$, and the matrix $M$ is

$$M = D(vec(\Sigma), vec(\Lambda))(\Psi + \Phi)D(vec(\Sigma), vec(\Lambda))', \tag{3.8b}$$

where $D(vec(\Sigma), vec(\Lambda))'$ is a $p^2 \times 2p^2$ matrix defined in (A.8) and $\Psi$ and $\Phi$ are $2p^2 \times 2p^2$ matrices presented respectively in (A.4) and (A.5). The obvious course of action in (3.8a)-(3.8b) is to replace the parameters by consistent estimates. In this fashion, $C_2$ is replaced by the $p \times (p-k)$ matrix $\hat{C}_2$ of eigenvectors relative to the $p-k$ smallest eigenvalues of the matrix $I_p - S^{-1/2}\hat{\Lambda}_n S^{-1/2}$, $\Sigma$ is replaced by $S$, and $\Lambda$ by $\hat{\Lambda}_n$. Some technical problems appear when building consistent estimates for the matrices $\Psi$ and $\Phi$. Approximate estimates $\hat{\Psi}$ and $\hat{\Phi}$ are given in equations (B.5) and (B.6) of appendix B, respectively. Finally,

$$\hat{\sigma}_k^2 = \hat{u}_k' \widehat{M} \hat{u}_k, \tag{3.9}$$

where $\hat{u}_k = vec(\hat{C}_2 \hat{C}_2')$ and $\widehat{M} = D(vec(S), vec(\hat{\Lambda}_n))(\hat{\Psi} + \hat{\Phi})D(vec(S), vec(\hat{\Lambda}_n))'$.

### 3.4 Final remarks

Let $\hat{\beta}_1 \geq \hat{\beta}_2 \geq \ ... \ \geq \hat{\beta}_p$ be the ordered eigenvalues of the matrix $S^{-1/2}\hat{\Lambda}_n S^{-1/2}$. It is clear that an alternative form for (3.6) is

$$R_k = n^{1/2}[\sum_{j=1}^{p-k} \hat{\beta}_j - (p-k)] \xrightarrow{D} N(0, \sigma_k^2). \tag{3.10}$$

Dimension can be then analyzed on the basis of the values of the statistics $R_k/\widehat{\sigma}_k$. On the other hand, the total variation matrix $T = \sum_{i=1}^{n}(x_i - \overline{x})(x_i - \overline{x})'$ can be decomposed

$$T = \sum_{h=1}^{H}\sum_{j=1}^{c}[x_{(h,j)} - \widehat{\mu}_{(h)}][x_{(h,j)} - \widehat{\mu}_{(h)}]' + \sum_{h=1}^{H} c\,[\widehat{\mu}_{(h)} - \overline{x}][\widehat{\mu}_{(h)} - \overline{x}]', \tag{3.11}$$

where the same caveat on notation made after (3.2) applies. Since $S = T/n$, dividing (3.11) by $n$ and pre and post multiplying by $S^{-1/2}$ leads to

$$I_p = \frac{c-1}{c} S^{-1/2}\widehat{\Lambda}_n S^{-1/2} + \widehat{V}_n, \tag{3.12}$$

where $\widehat{V}_n$ is exactly the $p \times p$ weighted covariance matrix (2.3) for the case of slicing using the same number of observations per slice. From (3.12),

$$I_p - S^{-1/2}\widehat{\Lambda}_n S^{-1/2} = \widehat{V}_n - \frac{1}{c} S^{-1/2}\widehat{\Lambda}_n S^{-1/2}, \tag{3.13}$$

so the $\widehat{\alpha}_j$ are the eigenvalues of the matrix $\widehat{V}_n - \frac{1}{c} S^{-1/2}\widehat{\Lambda}_n S^{-1/2}$. The statistics $S_k$ of (3.6) and $L_k$ of (2.4) have then similar structures. Relation (3.13) will be explored further in subsection 4.3 below.

## 4. SIMULATIONS, EXAMPLES AND COMPARISONS

This section studies empirical aspects of the test of dimensionality constructed in section 3. Simulation techniques are used to assess the accuracy of the approximation

$$G_k = S_k/\widehat{\sigma}_k \overset{D}{\sim} N(0,1),$$

in different setups. The role as a tuning constant of the number $c$ of observations in each slice is analyzed as well. The dimensionality of the air quality data of subsection 2.2 is tested using the methodology based on $G_k$. Finally, a Monte Carlo study compares $G_k$ with the statistic

$$L_k = n \sum_{i=k+1}^{p} \widehat{\lambda}_i,$$

of Li (1991) introduced in (2.4).

### 4.1 Distribution of $G_k$ and the role of $c$

The case $k = 1$ is studied. For samples sizes $n = 100$, 150 and 200, $N = 500$ independent replicates of a structure of the form

$$y = \exp[-(x_1 + x_2 + 2x_3)] + .5\epsilon, \tag{4.1}$$

are generated. The dimension $p$ of the regressors $\mathbf{x}$ is taken to be either $p = 3$ or 5. The error $\epsilon$ in (4.1) is $N(0,1)$ and is independently distributed from the predictors. For generating $\mathbf{x}$, three different models are used:

- **Model A**: $\mathbf{N}_p(0, \Sigma)$, where $\Sigma = diag(4, 9, 4, 2, 2)$;
- **Model B**: An elliptical model obtained dividing a $N_p(0, \Xi)$ random vector, where $\Xi = diag(2, 3, 2, 2, 2)$, by the squared root of an independent $\chi^2_{18}$ random variable divided by its degrees of freedom;
- **Model C**: A general distribution satisfying condition (2.1) defined as follows. Suppose that, as in (1.1), the true regression relation is $y = f(\beta_1'\mathbf{x}, \ ..., \beta_k'\mathbf{x}, \epsilon)$. Putting $B = (\beta_1, \ ..., \ \beta_k)$, generate two independent random vectors $\mathbf{v}$ and $\mathbf{w}$ of appropriate dimensions and define

$$\mathbf{x} = C\mathbf{v} + B(B'B)^{-1}\mathbf{w}, \tag{4.2}$$

where $C$ is a $p \times (p-k)$ matrix of rank $p-k$ such that $C'B = 0$. From (4.2), $B'\mathbf{x} = \mathbf{w}$ and by independence of $\mathbf{v}$ and $\mathbf{w}$,

$$E(\mathbf{x} \mid B'\mathbf{x}) = E(\mathbf{x} \mid \mathbf{w}) = CE(\mathbf{v}) + B(B'B)^{-1}\mathbf{w} = CE(\mathbf{v}) + B(B'B)^{-1}B'\mathbf{x}$$

so condition (2.1) is obviously satisfied. Since $\mathbf{v}$ and $\mathbf{w}$ can have arbitrary distributions, the distribution of $\mathbf{x}$ in (4.2) can be far from being elliptical. For example, in this simulation exercise, for the case $p = 3$, the elements in (4.2) are

$$C = \begin{pmatrix} 1 & 1 \\ -1 & 1 \\ 0 & -1 \end{pmatrix} \quad , \quad B = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix},$$

$\mathbf{v} = (v_1, v_2)'$ is formed by two *i.i.d.* $U(-4, 4)$ random variables, and $\mathbf{w} = w_1$ is $N(0, 1)$.

Model A is then the pure multinormal case, model B is an elliptically symmetric distribution longer tailed than the normal, and model C is just a distribution satisfying condition (2.1) but not belonging to any specified family of distributions. Finally, the scale estimate $\widehat{\sigma}_1$ is constructed using equation (3.9) and companion expressions and recommendations of appendix B.

*Figure* 2

*Table* 2

Figure 2 displays a set of histograms of the $N = 500$ replications corresponding to different models for generating $\mathbf{x}$ and different choices of $c$. The sample size is $n = 200$ for both models **A** and **C**, and $n = 150$ for model **B**. In all the plots, the dimension of the predictors is $p = 3$. Table 2 gives the empirical tail probabilities for each one of the histograms in figure 2 corresponding to the upper $(1 - \alpha) \times 100\%$ quantiles $z_\alpha$ of a $N(0, 1)$ distribution for values of $\alpha = .1,\ .05$, and $.01$. Notice that although all the histograms in figure 2 are bell shaped, the asymptotic $N(0, 1)$ approximation for the null distribution of the statistic $G_1$ is quite reliable in the tails for small values of $c$. However, the approximation deteriorates when $c$ increases. This is in agreement with the asymptotic theory presented in section 3, that implicitly assumes that the number of slices $H = H_n \sim n/c$ is "large" or, equivalently, that the number $c$ of responses in each slice is small.

### 4.2 Air quality data (continued)

The dimension of the air quality data of example 2.2 is analyzed using the sequential testing methodology based on the statistics $G_k$ introduced in section 3. Table 3 displays, for different choices of $c$, the values of the partial sums of eigenvalues $S_k$, the estimates $\widehat{\sigma}_k$. the statistics $G_k$ and the associated p-values using the $N(0, 1)$

distribution. Observe that, regardless of the value of $c$, the dimension of the regression for this set of data is declared always at least one. This conclusion is in contrast with the results from table 1, where the behaviour of Li's statistic is clearly affected by the form the $H$ slices are chosen.

*Table* 3

### 4.3 Power of $G_k$ and comparison with $L_k$

The case $k = 2$ is considered. For sample sizes $n = 100$, 200, and 300, $N = 1000$ independent replications of a regression

$$y = (4 + x_1)(2 + x_2 + x_3) + .5\epsilon, \tag{4.3}$$

are generated. The dimension of the regressors $\mathbf{x}$ is $p = 3$. As usual, $\epsilon$ is $N(0,1)$ and is independent of the regressors. For generating $\mathbf{x}$, three different models are considered:

- **Model A**: $\mathbf{N}_p(0, \Sigma)$, where $\Sigma = diag(2, 2, 2)$;
- **Model B**: An elliptical model obtained dividing a $N_p(0, \Xi)$ random vector, where $\Xi = diag(2, 2, 2)$, by the squared root of an independent $\chi^2_{16}$ random variable divided by its degrees of freedom;
- **Model C**: A distribution of the form (4.2), where

$$C = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix},$$

  $\mathbf{v} = v_1 \sim U(-4, 4)$, and $\mathbf{w} = (w_1, w_2)'$ is formed by two independent components such that $w_1$ is a mixture $.5N(0, \sigma^2 = 4) + .5N(0, \sigma^2 = 16)$ of two normals, and $w_2$ is $U(-4, 4)$.

For a nominal size $\alpha = .05$, and for each selection of the type of model and sample size $n$, the null hypotheses $H_k : rank(V) = k$, $k = 0$, 1, 2, are tested sequentially

using both the statistics $G_k$ and $L_k$ for several values of the slicing tuning constant $c$. The statistics $L_k$ are computed using the eigenvalues of the matrix $\widehat{V}_n$ introduced in (3.12). Results are presented in table 4. The entries in the rows corresponding to $k = 0$ and $k = 1$ are the proportions of times the nulls $H_0$ and $H_1$ are rejected, and can be then considered as empirical estimates of the power of the corresponding tests of dimension when the null $H_2$ is the truth. In this same vein, the entries in the rows labeled with $k = 2$ are empirical estimates of the actual size of the test.

From table 4, a general pattern emerges. For increasing sample sizes $n$ and small values of $c$, and regardless of the model that generates $\mathbf{x}$, the size of the rejection rule $G_2 > 1.645$ is close to the nominal size $\alpha = .05$. Moreover, for $k = 0$ and $k = 1$, the rules $G_k > 1.645$ have reasonable large power. As a comparison, $L_k$ behaves in an opposite fashion. For large values of $c$, the rule $L_2 > \chi^2_{(H-3),.05}$, where $H = H_n = [n/c] + 1$ and $\chi^2_{(H-3),.05}$ is the 95% upper quantile of a chi squared distribution with $H - 3$ degrees of freedom, has a size close to .05. For $k = 0$ and $k = 1$, the rules $L_k > \chi^2_{(3-k)(H-k-1),.05}$ have large power.

The reason for this phenomenon is not easy to find. Observe first that, in this simulation, the number $H$ of slices is a function of both the sample size $n$ and the number $c$ of observations that are included in each slice so, in contrast with the usual approach of keeping the slices fixed, the slicing criterion is data adaptive. This explains why in table 4, even when the normal model A is used for generating $\mathbf{x}$, the asymptotic chi squared approximation for the null distribution of $L_2$ fails to provide the correct sizes for small values of $c$. A possible heuristic argument for explaining the observed rows and columns in table 4 can be given as follows. According to relation (3.13) of subsection 3.4, the matrices $\widehat{V}_n$ and $I_3 - S^{-1/2}\widehat{\Lambda}_n S^{-1/2}$ are related, so the associated eigenvalues $(\widehat{\lambda}_1, \widehat{\lambda}_2, \widehat{\lambda}_3)$ of $\widehat{V}_n$ and $(\widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{\alpha}_3)$ of $I_3 - S^{-1/2}\widehat{\Lambda}_n S^{-1/2}$, are expected to be related as well. In fact, for $c$ large, the approximations

$$\widehat{\lambda}_3 \cong \widehat{\alpha}_3 + \frac{1}{c}, \tag{4.4}$$

and

$$\frac{\sqrt{n}\hat{\sigma}_2}{\sqrt{2H}} \cong 1, \tag{4.5}$$

were observed to hold in the setup of this simulation. From (4.4) and (4.5),

$$\begin{aligned} L_2 &= n\hat{\lambda}_3 \cong \sqrt{n}S_2 + \frac{n}{c} = \sqrt{n}\hat{\sigma}_2 G_2 + \frac{n}{c} \\ &\overset{D}{\cong} \sqrt{n}\hat{\sigma}_2 \frac{\chi^2_H - H}{\sqrt{2H}} + \frac{n}{c} \overset{D}{\cong} \chi^2_H, \end{aligned} \tag{4.6}$$

so it is reasonable to expect, using (4.6), some closeness of the distribution of $L_2$ to the appropriate chi squared distribution.

Other criteria could have been included in this comparative study, for example the statistics $T^{(i)}_{j,k}$ proposed by Schott (1994) for the case when $\mathbf{x}$ has an elliptically symmetric distribution. However, the asymptotic theory in Schott's criteria assumes that the number $c$ of observations in each slice grows to infinity. Notice that, in table 4, $c$ is a small number so clearly the methods of Schott (1994) and $G_k$ of subsection 3.2 are of different nature. For a detailed study on the behaviour of $T^{(i)}_{j,k}$ see section 4 in Schott (1994, p. 144 and 145).

## 5. CONCLUSIONS

This paper presents a method for testing for dimension in a general regression model as (1.1) that, unlike previously proposed methods, does not depend on specific assumptions on the distribution of the regressors. The method is numerically feasible and has a satisfactory behavior both in simulations and with real data.

A comparison with the test $L_k$ of dimension of Li (1991) is also given. Although the results obtained in table 4 suggest some kind of robustness of Li's (1991) criterion for finding dimension, the argument developed in (4.4), (4.5) and (4.6) indicates that the results found there are strongly dependent on the design of the simulation and do not need to hold in more general settings. Therefore, on the light of the asymptotic theory of section 3, $G_k$ with an associated low value of $c$ seems to be, as a general rule, a reliable option.

$G_k$ is based on a clearly defined criterion for grouping the data, including the same number $c$ of observations in each slice. Although Li (1991) comments on this issue in his remark 5.3, there is not such a parallel for $L_k$. In fact, as the analysis of the air quality data given in subsections 2.2 and 4.2 shows, the criterion used for choosing the slices when applying $L_k$ can lead to different conclusions on the dimension of a regression relating a response $y$ with a set of regressors $\mathbf{x}$.

Theorem 1 is motivated by the convergence (3.3). An alternative approach has been proposed by Zhu and Fang (1996) where they establish convergence of an estimate of $\Lambda$ based on kernel density estimation. Details and comparison with the methodology of section 3.2 are under investigation.

## APPENDIX A: PROOF OF THEOREM 1

*First part.* Put $m(y) = E(\mathbf{x} \mid y)$ and $\mathbf{u} = \mathbf{x} - m(y)$. Consider the $p \times 2p$ matrix

$$n^{1/2}[(S, \widehat{\Lambda}_n) - (\Sigma, \Lambda)], \tag{A.1}$$

where $S = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(x_i - \overline{x})'$ and $\widehat{\Lambda}_n$ is as defined in (3.2). One has $n^{1/2}[(S, \widehat{\Lambda}_n) - (\Sigma, \Lambda)] = n^{1/2}(S - \Sigma, \widehat{\Lambda}_n - \Lambda) =$

$$n^{1/2}[\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)(x_i - \mu)' - \Sigma\,,\, (\frac{1}{n}\sum_{i=1}^{n}u_i u_i' - \Lambda) + \frac{1}{n(c-1)}\sum_{h=1}^{H}Q_h] + o_p(1), \tag{A.2}$$

where $u_i = x_i - m(y_i)$, $Q_h = \sum\sum_{1 \le j \ne l \le c} u_{(h,j)}u_{(h,l)}'$ and $u_{(h,j)} = x_{(h,j)} - m(y_{(h,j)})$, where $x_{(h,j)}$ is the concomitant value of the $jth$ order statistic in the $hth$ slice. Equation (A.2) follows from expressions (3.7) and (3.8) in Zhu and Ng (1995, p. 733 and 734). The vectorization of the matrix of (A.1) is

$$n^{1/2}[\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} vec[(x_i - \mu)(x_i - \mu)'] \\ vec(u_i u_i') \end{pmatrix} - \begin{pmatrix} vec(\Sigma) \\ vec(\Lambda) \end{pmatrix}] \tag{A.3}$$

$$+ n^{1/2}[\frac{1}{n}\sum_{h=1}^{H}\begin{pmatrix} 0 \\ vec(Q_h)/(c-1) \end{pmatrix}] + o_p(1).$$

Moreover, $vec[(x_i-\mu)(x_i-\mu)'] = vec[u_iu_i'+u_i(m(y_i)-\mu)'+(m(y_i)-\mu)u_i'+(m(y_i)-\mu)(m(y_i)-\mu)']$. The concomitants $u_{(i)}$ are conditionally independent given the order statistics $y_{(1)}$, ..., $y_{(n)}$ (Yang 1977). Arguing as in theorem 3.2 in Hsing and Carroll (1992, p. 1045) or theorem 1 in Zhu and Ng (1995, p. 730) the limit distribution in (A.3) can be established as the convolution of two normal distributions with zero mean. The first one has variance-covariance matrix

$$\Psi = \begin{pmatrix} V(vec[(\mathbf{x}-\mu)(\mathbf{x}-\mu)']) & C(vec[(\mathbf{x}-\mu)(\mathbf{x}-\mu)'], vec[\mathbf{uu}']) \\ C(vec[\mathbf{uu}'], vec[(\mathbf{x}-\mu)(\mathbf{x}-\mu)']) & V(vec[\mathbf{uu}']) \end{pmatrix}, \tag{A.4}$$

and the second,

$$\Phi = \begin{pmatrix} 0 & 0 \\ 0 & \frac{2}{c-1} E[vec(V(y))\, vec(V(y))'] \end{pmatrix}, \tag{A.5}$$

where $V(y) = V(\mathbf{u}\,|y)$. As a conclusion,

$$n^{1/2}[vec(S, \hat{\Lambda}_n) - vec(\Sigma, \Lambda)] = n^{1/2}[\begin{pmatrix} vec(S) \\ vec(\hat{\Lambda}_n) \end{pmatrix} - \begin{pmatrix} vec(\Sigma) \\ vec(\Lambda) \end{pmatrix}] \xrightarrow{D} \mathbf{P}, \tag{A.6}$$

where $\mathbf{P} \sim \mathbf{N}_{p^2}(0, \Psi + \Phi)$. ■

*Second part.* For a differentiable function $\mathbf{g} : \mathbf{R}^q \to \mathbf{R}^s$ use the notation $\partial \mathbf{g}/\partial \mathbf{t} = (\partial g_i/\partial t_j)$ $_{i=1, ..., s;\, j=1, ..., q}$ for the $s \times q$ matrix of partial derivatives. By the delta method, the asymptotic distribution of

$$n^{1/2} vec(S^{-1/2}\hat{\Lambda}_n S^{-1/2} - \Sigma^{-1/2}\Lambda\Sigma^{-1/2}) = n^{1/2}[\mathbf{g}(vec(S), vec(\hat{\Lambda}_n)) - \mathbf{g}(vec(\Sigma), vec(\Lambda))],$$

where $\mathbf{g}(vec(\Sigma), vec(\Lambda)) = vec(\Sigma^{-1/2}\Lambda\Sigma^{-1/2})$, is the distribution of the $p^2 \times 1$ random vector

$$\mathbf{Q} = D(vec(\Sigma), vec(\Lambda))\mathbf{P} \tag{A.7}$$

where $\mathbf{P}$ is as introduced in (A.6), and

$$D(vec(\Sigma), vec(\Lambda)) = (\partial vec(\Sigma^{-1/2}\Lambda\Sigma^{-1/2})/\partial vec(\Sigma), \partial vec(\Sigma^{-1/2}\Lambda\Sigma^{-1/2})/\partial vec(\Lambda)), \tag{A.8}$$

is a $p^2 \times 2p^2$ matrix. By the chain rule,

$$\begin{aligned} \partial vec(\Sigma^{-1/2}\Lambda\Sigma^{-1/2})/\partial vec(\Sigma) \;=\; & [\partial vec(\Sigma^{-1/2}\Lambda\Sigma^{-1/2})/\partial vec(\Sigma^{-1/2})] \\ & [\partial vec(\Sigma^{-1})/\partial vec(\Sigma^{-1/2})]^{-1}[\partial vec(\Sigma^{-1})/\partial vec(\Sigma)] \end{aligned} \tag{A.9}$$

Using expressions (4) and (5) in Fang and Zhang (1990, p. 19),

$$\partial vec(\Sigma^{-1/2}\Lambda\Sigma^{-1/2})/\partial vec(\Sigma^{-1/2}) = (\Sigma^{-1/2}\Lambda \otimes I_p)K_p + (I_p \otimes \Sigma^{-1/2}\Lambda), \tag{A.9a}$$

where $K_p = \sum_{i=1}^{p}\sum_{j=1}^{p}(e_ie_j')\otimes(e_je_i')$ is the permutation matrix of order $p$, being $e_i$ the $ith$ canonical vector of $\mathbf{R}^p$,

$$\partial vec(\Sigma^{-1})/\partial vec(\Sigma^{-1/2}) = (\Sigma^{-1/2} \otimes I_p)K_p + (I_p \otimes \Sigma^{-1/2}), \tag{A.9b}$$

and

$$\partial vec(\Sigma^{-1})/\partial vec(\Sigma) = -(\Sigma^{-1} \otimes \Sigma^{-1}). \tag{A.9c}$$

Finally, by expression (8) in Fang and Zhang (1990, sec. 1.4.2, p. 12), $vec(\Sigma^{-1/2}\Lambda\Sigma^{-1/2}) = (\Sigma^{-1/2} \otimes \Sigma^{-1/2})vec(\Lambda)$, where $\otimes$ denotes Kronecker product. Therefore,

$$\partial vec(\Sigma^{-1/2}\Lambda\Sigma^{-1/2})/\partial vec(\Lambda) = \Sigma^{-1/2} \otimes \Sigma^{-1/2}. \tag{A.10}$$

■

*Third part.* By the first two parts,

$$n^{1/2}[(I_p - S^{-1/2}\hat{\Lambda}_n S^{-1/2}) - V] \xrightarrow{D} \mathbf{R},$$

where $\mathbf{R}$ is a $p \times p$ random symmetric matrix such that $vec(\mathbf{R}) = \mathbf{Q}$, where $\mathbf{Q}$ is as introduced in (A.7). Write

$$V = CDC', \tag{A.11}$$

where $D = diag(\lambda_1, \ldots, \lambda_p)$ is the $p \times p$ diagonal matrix of ordered eigenvalues of $V$ and $C = (\gamma_1, \ldots, \gamma_p)$ is a $p \times p$ orthogonal matrix of eigenvectors. If $rank(V) = k$, then $\lambda_1 \geq \ldots \geq \lambda_k > 0 = \lambda_{k+1} = \ldots = \lambda_p$. From theorem 3.2 in Eaton and Tyler (1991, p. 265),

$$n^{1/2} \begin{pmatrix} \widehat{\alpha}_{k+1} \\ \widehat{\alpha}_{k+2} \\ \vdots \\ \widehat{\alpha}_p \end{pmatrix} \xrightarrow{D} \varphi(C_2'\mathbf{R}C_2), \tag{A.12}$$

where $\varphi(C_2'\mathbf{R}C_2)$ is the vector of ordered eigenvalues of $C_2'\mathbf{R}C_2$, being $C_2 = (\gamma_{k+1}, \ldots, \gamma_p)$ the $p \times (p-k)$ block of $C$ corresponding to the $(p-k)$ null eigenvalues in $D$. From (A.12) and identity (6) in Fang and Zhang (1990, sec. 1.4.1, p. 12),

$$n^{1/2} \sum_{j=k+1}^{p} \widehat{\alpha}_j \xrightarrow{D} tr(C_2'\mathbf{R}C_2) = tr(C_2C_2'\mathbf{R}) = (vec(C_2C_2'))'vec(\mathbf{R}). \tag{A.13}$$

Since $vec(\mathbf{R}) = \mathbf{Q} = D(vec(\Sigma), vec(\Lambda))\mathbf{P}$, the limit distribution in (A.13) is $N(0, \sigma_k^2)$ where

$$\sigma_k^2 = (vec(C_2C_2'))'D(vec(\Sigma), vec(\Lambda))(\Psi + \Phi)D(vec(\Sigma), vec(\Lambda))'vec(C_2C_2'), \tag{A.14}$$

and the matrices $\Psi$, $\Phi$ and $D(vec(\Sigma), vec(\Lambda))$ are as defined in (A.4), (A.5) and (A.8), respectively. Observe that the dependence of $\sigma_k^2$ on $k$ is through the number of eigenvectors discarded to get the submatrix $C_2$ of $C$. ■

## APPENDIX B: ESTIMATION OF $\Psi$ AND $\Phi$

Recall that for every $p \times 1$ vector $a$, $vec(aa') = a \bigotimes a$. The matrix $\Psi$ of (A.4) can be written then in the form

$$\Psi = \begin{pmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{pmatrix},$$

where $\Psi_{11} = V[(\mathbf{x}-\mu) \bigotimes (\mathbf{x}-\mu)]$, $\Psi_{12} = C[(\mathbf{x}-\mu) \bigotimes (\mathbf{x}-\mu), \mathbf{u} \bigotimes \mathbf{u}]$, $\Psi_{21} = \Psi_{12}'$ and $\Psi_{22} = V[\mathbf{u} \bigotimes \mathbf{u}]$, where $\mathbf{u} = \mathbf{x} - E(\mathbf{x} \mid y)$. Put, for $i = 1, \ldots, n$,

$$\Omega_i = (x_i - \overline{x}) \bigotimes (x_i - \overline{x}), \tag{B.1}$$

and

$$\Pi_i = [x_i - \widehat{m}_n(y_i)] \bigotimes [x_i - \widehat{m}_n(y_i)], \tag{B.2}$$

where $x_i$ is the $ith$ regressor point, $\overline{x}$ is the sample mean vector, and $\widehat{m}_n(y_i)$ is a nonparametric estimate of the inverse regression curve $m(y) = E(\mathbf{x} \mid y)$ evaluated at the $ith$ response. The matrix $\Psi$ of (A.4) is estimated using

$$\widehat{\Psi} = \begin{pmatrix} \widehat{\Psi}_{11} & \widehat{\Psi}_{12} \\ \widehat{\Psi}_{21} & \widehat{\Psi}_{22} \end{pmatrix}, \tag{B.3}$$

where

$$\widehat{\Psi}_{11} = \frac{1}{n}\sum_{i=1}^{n} (\Omega_i - \overline{\Omega})(\Omega_i - \overline{\Omega})', \tag{B.3a}$$

$$\widehat{\Psi}_{12} = \frac{1}{n}\sum_{i=1}^{n} (\Omega_i - \overline{\Omega})(\Pi_i - \overline{\Pi})', \tag{B.3b}$$

$$\widehat{\Psi}_{21} = \widehat{\Psi}_{12}', \tag{B.3c}$$

and

$$\widehat{\Psi}_{22} = \frac{1}{n}\sum_{i=1}^{n}(\Pi_i - \overline{\Pi})(\Pi_i - \overline{\Pi})' , \tag{B.3d}$$

being $\overline{\Omega} = \frac{1}{n}\sum_{i=1}^{n}\Omega_i$, and $\overline{\Pi} = \frac{1}{n}\sum_{i=1}^{n}\Pi_i$.

In this paper, $\widehat{m}_n(y_i)$ is taken as a simple slice average smoother (Cook and Weisberg, 1994, p. 32) constructed averaging the regressors corresponding to the responses included in a fixed inreval of constant length $l$. For example, in the simulation in subsection 4.1, $l = .3$ while, in subsection 4.3, $l = 1.5$. In the air quality data studied in subsection 4.2, there are repeated responses and $\widehat{m}_n(y_i)$ was constructed averaging, for each value of $y$, the corresponding values of the predictors. This naive method for getting $\widehat{m}_n(y_i)$ was later perfected using the locally weighted scatterplot smoother (*"lowess"*) proposed by Cleveland (1979). As in Cook and Weisberg (1994), the lowess smoother was computed using step 1 of the algorithm 6.1.1 in Härdle (1990, p. 192). However, no real significant improvement was observed in the estimation of the matrix $\Psi$. In the light of these findings, a general recommendation in applications is to use slice averages smoothers for estimating $E(\mathbf{x}\,|y)$.

On the other hand, since $\mathbf{u} = \mathbf{x} - E(\mathbf{x}\,|y)$ one has $V(y) = V(\mathbf{u}\,|y) = V(\mathbf{x} \mid y)$. The generic element of the matrix $E[vec(V(y))\,vec(V(y))']$ is of the form

$$E\{cov[(x_j, x_k)\,|y]\,cov[(x_r, x_s)\,|y]\} , \tag{B.4a}$$

for $j,\ k,\ r,\ s = 1,\ ...,\ p$. (B.4a) can be estimated using

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{cov}[(x_j, x_k)\,|y_i]\,\widehat{cov}[(x_r, x_s)\,|y_i] , \tag{B.4b}$$

where

$$\widehat{cov}[(x_j, x_k)\,|y_i] = \widehat{E}(x_j x_k\,|y_i) - \widehat{E}(x_j\,|y_i)\widehat{E}(x_k\,|y_i), \tag{B.4c}$$

and $\widehat{E}(x_j x_k\,|y_i)$ and $\widehat{E}(x_j\,|y_i)$ are nonparametric estimates of the regression curves $E(x_j x_k\,|y_i)$ and $E(x_j\,|y_i)$ respectively, obtained using the corresponding smoothers. Observe that $\widehat{E}(x_j\,|y_i)$ is the $jth$ coordinate of $\widehat{m}_n(y_i)$. The matrix $\Phi$ is estimated using

$$\widehat{\Phi} = \frac{2}{c-1}\begin{pmatrix} 0 & 0 \\ 0 & \widehat{\Phi}_{22} \end{pmatrix}, \tag{B.4}$$

where $\widehat{\Phi}_{22}$ is a $p^2 \times p^2$ matrix obtained grouping together the expressions (B.4a), (B.4b), and (B.4c) above.

## REFERENCES


[1] Cleveland, W. (1979), "Robust locally weighted regression and smoothing scatterplots." *Journal of the American Statistical Association,* **74**, 829-836.

[2] Cook, R. D. (1994), "On the Interpretation of Regression Plots," *Journal of the American Statistical Association,* **89**, 177-189.

[3] Cook, R. D. and Weisberg, S. (1991), Comment on Li (1991), *Journal of the American Statistical Association,* **86**, 328-332.

[4] Cook, R. D. and Wetzel, N. (1993), "Exploring Regression Structure with Graphics (with discussion)," *Test,* **2**, 33-100.

[5] Cook, R. D. and Weisberg, S. (1994), *An Introduction to Regression Graphics,* New York: John Wiley.

[6] Dawid, A. P. (1979), "Conditional Independence in Statistical Theory (with discussion)," *Journal of the Royal Statistical Society, Ser. B,* **41**, 1-31.

[7] Duan, N. and Li, K.C. (1991), "Slicing Regression: A Link-Free Regression Method," *The Annals of Statistics*, **19**, 505-530.

[8] Eaton, M. L. and Tyler, D. E. (1991), "On Wielandt's Inequality and its Application to the Asymptotic Distribution of the Eigenvalues of a Random Symmetric Matrix," *The Annals of Statistics*, **19**, 260-271.

[9] Fang, K. T. and Zhang, Y. T. (1990), *Generalized Multivariate Analysis,* New York: Springer Verlag.

[10] Härdle, W. (1990), *Applied Nonparametric Regression.* Cambridge: Cambridge University Press.

[11] Hsing, T. and Carroll, R. (1992), "An Asymptotic Theory for Sliced Inverse Regression," *The Annals of Statistics*, **20**, 1040-1061.

[12] Li, K. C. (1991), "Sliced Inverse Regression for Dimension Reduction (with discussion)," *Journal of the American Statistical Association,* **86**, 316-342.

[13] Li, K. C. (1991), "On Principal Hessian Directions for Data Visualization and Dimension Reduction: Another Application of Stein's Lemma," *Journal of the American Statistical Association,* **87**, 1025-1039.

[14] Li, K. C. and Duan, N. (1989), "Regression Analysis Under Link Violation," *The Annals of Statistics*, **17**, 1009-1052.

[15] Schott, J. R. (1994), "Determining the Dimensionality in Sliced Inverse Regression," *Journal of the American Statistical Association,* **89**, 141-148.

[16] Yang, S. S. (1977), "General Distribution Theory of the Concomitants of Order Statistics," *The Annals of Statistics*, **5**, 996-1002.

[17] Zhu, L. X. and Ng, K. W. (1995), "Asymptotics of Sliced Inverse Regression," *Statistica Sinica,* **5**, 727-736.

[18] Zhu, L. X. and Fang, K. T. (1996), "Asymptotics for Kernel Estimate of Sliced Inverse Regression," *The Annals of Statistics*, **24**, 1055-1068.

## CAPTIONS FOR FIGURES AND TABLES

**Figure 1.** Scatter plot of the air quality data obtained using the $R - code$.

**Table 1.** Analysis of the dimension of the air quality data based on the statistics $L_k$ of Li (1991).

**Figure 2.** Histograms of $N = 500$ independent replications of the distribution of $G_1$ under different models generating $\mathbf{x}$, different sample sizes $n$, and various choices of $c$.

**Table 2.** Empirical tail probabilities of the histograms of figure 2 corresponding to the .90, .95, and .99 upper quantiles of a $N(0,1)$ distribution.

**Table 3.** Sequential analysis of the dimension of the air quality data using the statistics $G_k$.

**Table 4.** Empirical size and power of $L_k$ and $G_k$ under the simulation design of subsection 4.3.

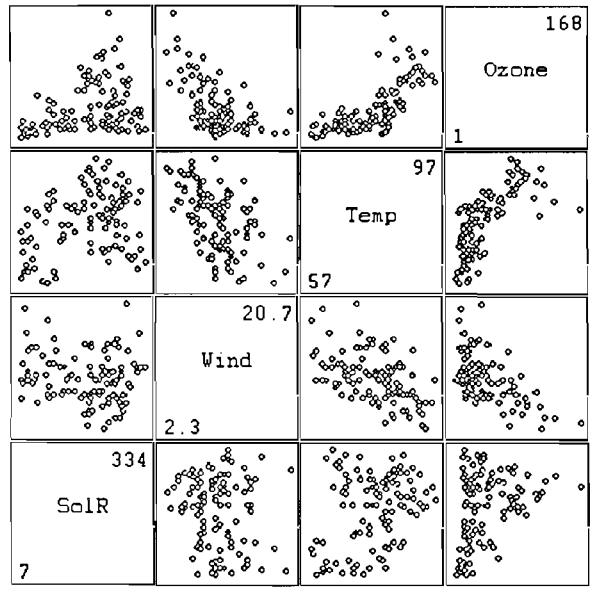


Figure 1

| H = 19 | k | $L_k$ | d.f | p − value |
|---|---|---|---|---|
| | 0 | 177.4 | 54 | .000 |
| | 1 | 33.96 | 34 | .470 |
| | 2 | 15.4 | 16 | .495 |

| H = 26 | k | $L_k$ | d.f | p − value |
|---|---|---|---|---|
| | 0 | 149.5 | 75 | .000 |
| | 1 | 64.73 | 48 | .054 |
| | 2 | 29.74 | 23 | .157 |

**Table 1**

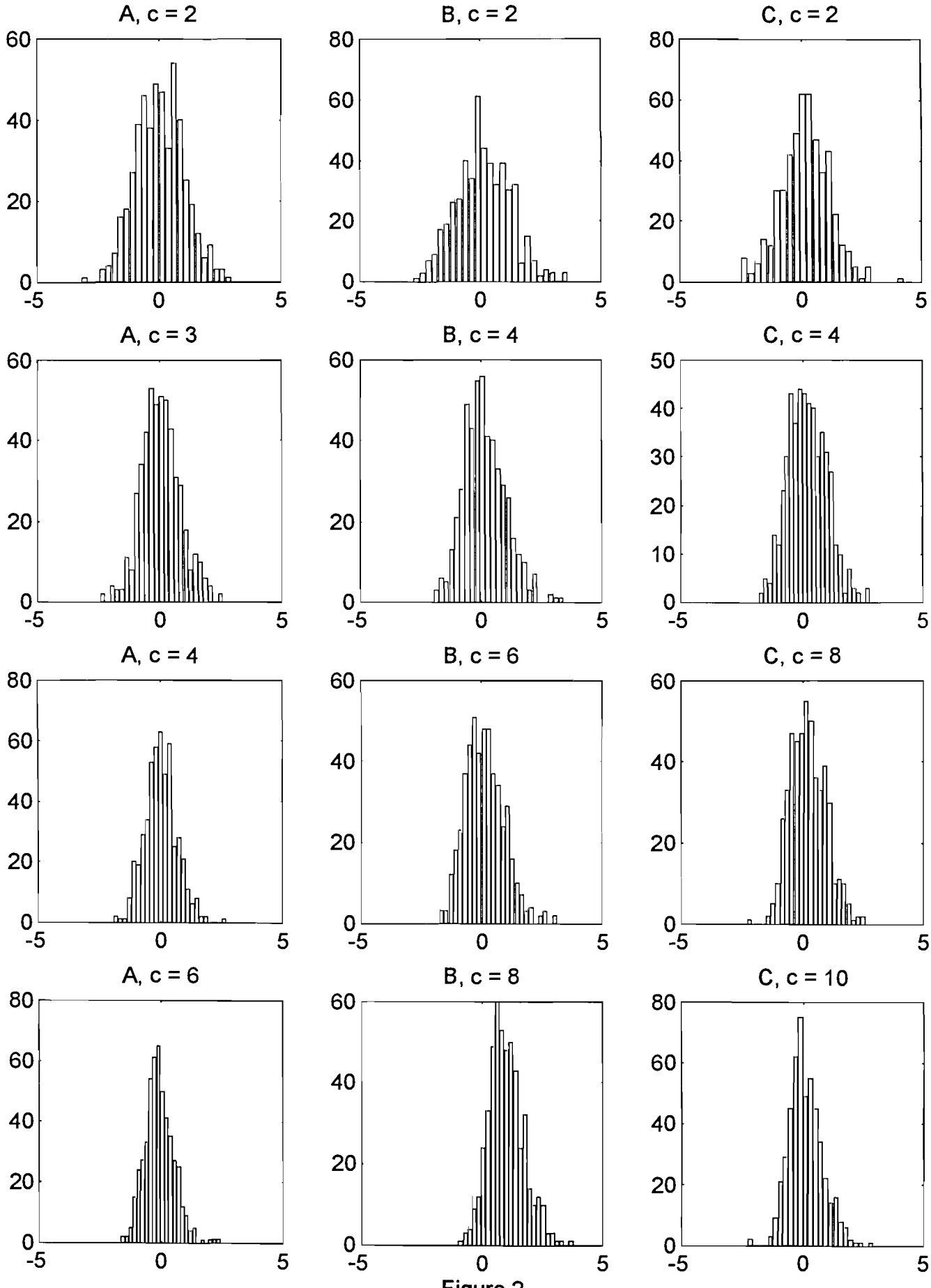


Figure 2

| Model A | c | $\alpha = .1$ | $\alpha = .05$ | $\alpha = .01$ |
|---|---|---|---|---|
| | 2 | .100 | .048 | .010 |
| | 3 | .064 | .032 | .004 |
| | 4 | .072 | .040 | .004 |
| | 5 | .034 | .008 | .002 |

| Model B | c | $\alpha = .1$ | $\alpha = .05$ | $\alpha = .01$ |
|---|---|---|---|---|
| | 2 | .146 | .080 | .048 |
| | 4 | .104 | .050 | .008 |
| | 6 | .070 | .034 | .014 |
| | 8 | .114 | .182 | .052 |

| Model C | c | $\alpha = .1$ | $\alpha = .05$ | $\alpha = .01$ |
|---|---|---|---|---|
| | 2 | .090 | .058 | .016 |
| | 4 | .084 | .038 | .006 |
| | 8 | .066 | .034 | .004 |
| | 10 | .064 | .022 | .004 |

**Table 2**

| $\mathbf{c}$ | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| $\mathbf{H}$ | 56 | 37 | 28 | 23 | 19 |
| $\mathbf{S_0}$ | 7.3890 | 9.6483 | 6.0609 | 7.0646 | 7.3151 |
| $\mathbf{S_1}$ | 0.8247 | 2.2805 | $-.2622$ | 0.1728 | 0.1990 |
| $\mathbf{S_2}$ | $-1.5303$ | 0.4019 | $-.7067$ | $-.1428$ | $-.4579$ |
| $\widehat{\sigma}_0$ | 2.5720 | 1.9847 | 1.7952 | 1.8703 | 1.6241 |
| $\widehat{\sigma}_1$ | 2.1875 | 1.7625 | 1.5769 | 1.5317 | 1.4398 |
| $\widehat{\sigma}_2$ | 1.3079 | 1.0222 | 1.0536 | 1.0159 | 0.8194 |
| $\mathbf{G_0}$ | 2.8729 | 4.8613 | 3.8774 | 3.1072 | 4.5041 |
| $\mathbf{G_1}$ | 0.3770 | 1.2239 | $-.1663$ | 0.1128 | 0.1382 |
| $\mathbf{G_2}$ | $-1.1701$ | 0.3932 | $-.6708$ | $-.1396$ | $-.5588$ |
| | | | $\mathbf{p-values}$ | | |
| $\mathbf{k=0}$ | .002 | .000 | .000 | .000 | .000 |
| $\mathbf{k=1}$ | .353 | .113 | .566 | .455 | .445 |
| $\mathbf{k=2}$ | .879 | .347 | .748 | .555 | .742 |

**Table 3**

| Model A | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | c | 2 | 3 | 4 | 5 |
| | n = 100 | k = 0 | $L_0$ | 0.989 | 1.000 | 1.000 | 1.000 |
| | | | $G_o$ | 0.985 | 1.000 | 0.999 | 1.000 |
| | | k = 1 | $L_1$ | 0.092 | 0.256 | 0.406 | 0.496 |
| | | | $G_1$ | 0.172 | 0.230 | 0.218 | 0.218 |
| | | k = 2 | $L_2$ | 0.001 | 0.007 | 0.021 | 0.020 |
| | | | $G_2$ | 0.016 | 0.020 | 0.011 | 0.009 |
| | | | c | 2 | 4 | 6 | 10 |
| | n = 200 | k = 0 | $L_0$ | 1.000 | 1.000 | 1.000 | 1.000 |
| | | | $G_0$ | 1.000 | 1.000 | 1.000 | 1.000 |
| | | k = 1 | $L_1$ | 0.198 | 0.650 | 0.827 | 0.920 |
| | | | $G_1$ | 0.333 | 0.487 | 0.567 | 0.588 |
| | | k = 2 | $L_2$ | 0.002 | 0.018 | 0.035 | 0.040 |
| | | | $G_2$ | 0.036 | 0.022 | 0.016 | 0.003 |
| | | | c | 2 | 4 | 10 | 15 |
| | n = 300 | k = 0 | $L_0$ | 1.000 | 1.000 | 1.000 | 1.000 |
| | | | $G_0$ | 1.000 | 1.000 | 1.000 | 1.000 |
| | | k = 1 | $L_1$ | 0.277 | 0.828 | 0.985 | 0.993 |
| | | | $G_1$ | 0.451 | 0.718 | 0.861 | 0.886 |
| | | k = 2 | $L_2$ | 0.001 | 0.023 | 0.034 | 0.040 |
| | | | $G_2$ | 0.046 | 0.036 | 0.008 | 0.003 |

| Model B | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | c | 2 | 3 | 4 | 5 |
| | n = 100 | k = 0 | $L_0$ | 0.963 | 1.000 | 1.000 | 1.000 |
| | | | $G_o$ | 0.967 | 0.995 | 0.998 | 0.998 |
| | | k = 1 | $L_1$ | 0.105 | 0.295 | 0.401 | 0.499 |
| | | | $G_1$ | 0.185 | 0.274 | 0.202 | 0.195 |
| | | k = 2 | $L_2$ | 0.002 | 0.006 | 0.014 | 0.015 |
| | | | $G_2$ | 0.023 | 0.017 | 0.009 | 0.006 |
| | | | c | 2 | 4 | 6 | 10 |
| | n = 200 | k = 0 | $L_0$ | 1.000 | 1.000 | 1.000 | 1.000 |
| | | | $G_0$ | 1.000 | 1.000 | 1.000 | 1.000 |
| | | k = 1 | $L_1$ | 0.162 | 0.620 | 0.773 | 0.887 |
| | | | $G_1$ | 0.290 | 0.444 | 0.525 | 0.549 |
| | | k = 2 | $L_2$ | 0.001 | 0.022 | 0.033 | 0.040 |
| | | | $G_2$ | 0.029 | 0.024 | 0.013 | 0.004 |
| | | | c | 2 | 4 | 10 | 15 |
| | n = 300 | k = 0 | $L_0$ | 1.000 | 1.000 | 1.000 | 1.000 |
| | | | $G_0$ | 1.000 | 1.000 | 1.000 | 1.000 |
| | | k = 1 | $L_1$ | 0.282 | 0.772 | 0.965 | 0.978 |
| | | | $G_1$ | 0.437 | 0.644 | 0.797 | 0.819 |
| | | k = 2 | $L_2$ | 0.005 | 0.018 | 0.026 | 0.037 |
| | | | $G_2$ | 0.049 | 0.027 | 0.004 | 0.001 |

| Model C | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | c | 2 | 3 | 4 | 5 |
| | n = 100 | k = 0 | $L_0$ | 0.916 | 0.992 | 0.997 | 1.000 |
| | | | $G_o$ | 0.941 | 0.986 | 0.993 | 0.998 |
| | | k = 1 | $L_1$ | 0.176 | 0.427 | 0.600 | 0.644 |
| | | | $G_1$ | 0.336 | 0.480 | 0.501 | 0.506 |
| | | k = 2 | $L_2$ | 0.006 | 0.015 | 0.022 | 0.023 |
| | | | $G_2$ | 0.038 | 0.045 | 0.028 | 0.020 |
| | | | c | 2 | 4 | 6 | 10 |
| | n = 200 | k = 0 | $L_0$ | 0.998 | 1.000 | 1.000 | 1.000 |
| | | | $G_0$ | 0.999 | 1.000 | 1.000 | 1.000 |
| | | k = 1 | $L_1$ | 0.330 | 0.744 | 0.855 | 0.932 |
| | | | $G_1$ | 0.493 | 0.675 | 0.737 | 0.773 |
| | | k = 2 | $L_2$ | 0.004 | 0.022 | 0.038 | 0.044 |
| | | | $G_2$ | 0.042 | 0.026 | 0.018 | 0.007 |
| | | | c | 2 | 4 | 10 | 15 |
| | n = 300 | k = 0 | $L_0$ | 1.000 | 1.000 | 1.000 | 1.000 |
| | | | $G_0$ | 1.000 | 1.000 | 1.000 | 1.000 |
| | | k = 1 | $L_1$ | 0.443 | 0.864 | 0.983 | 0.992 |
| | | | $G_1$ | 0.618 | 0.810 | 0.900 | 0.910 |
| | | k = 2 | $L_2$ | 0.005 | 0.033 | 0.048 | 0.046 |
| | | | $G_2$ | 0.053 | 0.039 | 0.010 | 0.006 |

Table 4